

L'évaluation thérapeutique et le niveau de preuve

D^r Michel Cucherat

Pharmacologie clinique, faculté de médecine Laennec, 69008 Lyon
mcu@upcl.univ-lyon1.fr

Objectifs

- Argumenter l'évaluation d'une thérapeutique et les niveaux de preuve des principales sources d'information.
- Argumenter une publication d'essai clinique ou une méta-analyse et critiquer une information thérapeutique.

POURQUOI ET COMMENT ÉVALUER UNE NOUVELLE THÉRAPEUTIQUE ?

L'évaluation thérapeutique est le processus de recherche qui a pour objectif d'apporter les preuves fiables et cliniquement pertinentes de l'efficacité et de la sécurité d'une nouvelle thérapeutique avant son utilisation en pratique médicale courante.

Il s'agit par exemple de démontrer qu'un nouvel hypocholestérolémiant réduit la fréquence des événements coronariens de manière substantielle sans entraîner d'effets délétères graves ou trop fréquents. À partir de ce moment, cette nouvelle molécule aura sa place en prévention dans la population étudiée par ses essais.

La notion de pertinence clinique est centrale, car il ne suffit pas de montrer qu'un traitement a une action biologique (pharmacologique dans le cas d'un médicament) mais il faut prouver qu'il améliore réellement l'état de santé. Par exemple, la finalité du traitement de l'ostéoporose post-ménopausique n'est pas d'augmenter la densité osseuse mais bien de prévenir les fractures (la diminution de la densité osseuse étant un facteur de risque). L'augmentation de la densité osseuse n'est que le moyen par lequel on espère apporter un bénéfice clinique au patient. Un nouveau traitement devra démontrer qu'il réduit la fréquence des fractures. Un effet sur la densité osseuse est insuffisamment pertinent sur plan clinique.

L'évaluation thérapeutique est réalisée par une série d'essais thérapeutiques qui sont regroupés à la fin dans une analyse unique : la méta-analyse.

Les principes de l'évaluation thérapeutique s'appliquent à tous les types d'intervention thérapeutique : médicament, dispositif, intervention chirurgicale, régime, stratégie de prise en charge,

etc., mais elle n'est vraiment systématique et complète que pour le médicament en raison des obligations réglementaires de l'autorisation de mise sur le marché.

Qu'est-ce qu'une preuve fiable ?

Un résultat d'évaluation thérapeutique est fiable quand il reflète la réalité avec exactitude, c'est-à-dire quand il ne conclut à l'efficacité que lorsque le traitement est effectivement efficace. Tous les types d'arguments pouvant être avancés pour justifier l'efficacité d'un traitement ne sont pas fiables. Certains peuvent être faussement positifs en raison de biais, et d'autres sont trop théoriques ou spéculatifs.

Il est impératif de disposer de preuves de fiabilité absolue afin d'éviter la diffusion de traitements sans effet, voire délétères, sur la base d'arguments faussement en faveur de leur efficacité. Ainsi certains antiarythmiques de classe 1c ont été utilisés pendant des années pour prévenir la mort subite après infarctus du myocarde sur la base d'arguments insuffisamment fiables et qui se sont avérés faussement positifs lorsqu'un essai thérapeutique randomisé a finalement montré un doublement de la mortalité avec ces traitements.

Les arguments insuffisamment fiables pour apporter la preuve de l'efficacité des traitements sont :

– les arguments basés sur les mécanismes d'action et la physiopathologie, car le déterminisme des maladies est extrêmement complexe et souvent partiellement connu. Un effet biologique n'implique pas forcément un bénéfice clinique. Ces raisonnements théoriques sont indispensables pour avoir l'idée de nouveaux

traitements à soumettre à l'évaluation, mais ne peuvent pas être des preuves à part entière ;

– l'expérience empirique du patient ou du médecin, en raison 1) de la part importante des phénomènes d'auto- et d'hétéro-suggestion dans la perception de l'état de santé des individus (effet placebo) ; 2) de l'évolution naturelle des maladies ; 3) de la petitesse des bénéfices escomptés : par exemple, la réduction relative de 20 % de la fréquence d'un événement entre 5 à 10 % au bout de 5 ans. L'observation sans point de comparaison d'un faible nombre de sujets ne permet pas de percevoir intuitivement un changement de probabilité aussi faible ;

– les études observationnelles qui comparent le devenir des patients traités dans la pratique courante avec le traitement d'intérêt au devenir de ceux qui ne reçoivent pas ce traitement. Cette comparaison est sujette au biais d'indication provenant du fait que les patients recevant le traitement ne sont certainement pas comparables aux autres : ils peuvent être par exemple plus gravement atteints ou, au contraire, à meilleur pronostic. Les techniques statistiques d'ajustement utilisées pour tenter de corriger ce problème ne garantissent pas la disparition de ce biais, comme le montre la comparaison des résultats des études observationnelles et des essais thérapeutiques dans les domaines où ces deux types sont disponibles.

Des solutions méthodologiques ont été développées pour dépasser les limites de ces différentes approches et obtenir un niveau de fiabilité maximal. L'ensemble de ces solutions constitue l'essai thérapeutique contrôlé, randomisé en double aveugle, qui apporte une protection contre tous les biais potentiels.

Niveau de preuve

Le niveau de preuve est un indicateur de l'aptitude d'un type d'argument qui a donné des preuves fiables de l'efficacité. Des échelles de niveau de preuve (tableau) ont été établies pour hiérarchiser les différents types d'études en fonction de leur niveau de fiabilité.

JUSTIFICATION DES PRINCIPES DE L'ESSAI CONTRÔLÉ RANDOMISÉ

Groupe contrôle

Dans sa forme la plus simple, un essai thérapeutique comprend deux groupes de patients. Un groupe, dit groupe expérimental ou groupe « traité », reçoit le traitement évalué tandis que l'autre groupe (dit groupe contrôle) ne le reçoit pas.

Un critère de jugement quantifie le devenir des patients dans l'étude : valeur moyenne d'un paramètre biologique, fréquence de survenue d'un événement clinique, du décès, etc. L'effet du traitement est mis en évidence par l'existence d'une différence statistiquement significative au niveau du critère de jugement entre le groupe traité et le groupe contrôle, objectivant un devenir plus favorable avec le traitement étudié que sans le traitement étudié. En fonction de la nature du critère de jugement, cet effet favorable du traitement pourra être, par exemple, une diminution de la fréquence des décès, ou l'obtention en moyenne d'une valeur du paramètre biologique d'intérêt plus favorable dans le groupe traité que dans le groupe contrôle.

Les patients du groupe contrôle peuvent recevoir soit un placebo (traitement sans effet mais indiscernable du traitement étudié, servant à assurer le double aveugle) soit un traitement actif. La plupart des essais versus placebo s'effectuent en fait « par-dessus » la stratégie de traitement validée. Tous les patients de l'essai reçoivent les traitements démontrés efficaces y compris ceux du groupe placebo. Ces essais répondent à la question : « Quel est le bénéfice supplémentaire qu'apporte le nouveau traitement lorsque celui-ci est ajouté à ceux déjà disponibles ? ». En effet, ces essais ne comparent pas le nouveau traitement au placebo mais l'ajout du nouveau traitement à la stratégie habituelle par rapport à la stratégie habituelle seule. De ce fait, ce type d'essai contre placebo ne pose pas de problème éthique.

QU'EST-CE QUI PEUT TOMBER À L'EXAMEN ?

Un essai thérapeutique contrôlé, randomisé a comparé une nouvelle statine au placebo en prévention primaire chez des sujets à faible risque ne présentant pas d'indication de statines.

L'essai a été réalisé en double aveugle à l'aide d'un placebo complètement similaire au traitement étudié. Le critère de jugement principal était l'épaisseur intima-média carotidienne mesurée par échographie au 24^e mois.

Au total, 672 patients ont été randomisés dans le groupe de la nouvelle statine et 645 dans le groupe placebo.

❶ Quel est le niveau de preuve de cette étude ?

❷ Au bout de 24 mois, l'épaisseur intima-média a été mesurée chez respectivement 511 et 587 patients des groupes traité et placebo. Dans le groupe statine, l'épaisseur intima-média avait diminué en moyenne de 0,0026 mm par an par rapport au début de l'étude tandis

qu'elle avait progressé en moyenne de 0,0256 mm par an dans le groupe placebo ($p = 0,02$ pour la différence entre les deux groupes). À partir de ces résultats, que peut-on conclure ?

❸ Ce résultat est-il à l'abri des biais ?

❹ Le critère de jugement utilisé permet-il de démontrer l'intérêt clinique de la nouvelle statine ?

Éléments de réponse dans un prochain numéro

Le groupe contrôle sert de point de comparaison, indispensable pour savoir ce qu'a effectivement apporté le traitement étudié dans la valeur du critère de jugement. Sans point de comparaison, comment savoir si la fréquence de décès observée dans un groupe de patients traités par le nouveau traitement est moindre que celle qui aurait été observée sans ce traitement ? Le groupe contrôle donne cette référence et permet de déduire l'effet propre du traitement par la différence entre les deux groupes.

Le groupe contrôle permet aussi de séparer l'effet propre du traitement et l'effet des facteurs confondants qui sont des facteurs entraînant une évolution favorable du critère de jugement indépendamment de l'effet du traitement étudié. Leur effet peut ainsi être confondu avec l'effet du traitement. Il s'agit de l'évolution naturelle de la maladie, de l'effet placebo, de la régression à la moyenne, etc.

Pour être valable, le groupe contrôle doit inclure des patients similaires aux patients traités, c'est-à-dire être contemporain, et pris en charge dans les mêmes contextes de soins. Le groupe contrôle doit être constitué de manière prospective en répartissant entre les deux groupes, au fil du temps et au niveau des mêmes structures de soins, les nouveaux patients présentant la pathologie étudiée. Les comparaisons historiques (avec des patients pris en charge avant l'apparition du nouveau traitement) ou géographiques (avec des patients pris en charge dans une autre structure n'ayant pas accès au nouveau traitement) ne sont pas satisfaisantes, car rien ne garantit que ces patients sont comparables à ceux traités.

Biais

En évaluation thérapeutique, il y a biais lorsque la différence finale observée entre le groupe traité et le groupe contrôle peut avoir été induite par un autre facteur que le traitement étudié. Une méthodologie d'étude présentant un biais donnera un résultat positif de manière systématique quel que soit en réalité l'effet du traitement.

POINTS FORTS

à retenir

- Le but de l'évaluation thérapeutique est de rechercher les preuves fiables et cliniquement pertinentes de l'efficacité et de la sécurité des nouveaux traitements.
- Seul l'essai contrôlé randomisé en double aveugle, analysé en intention de traiter avec remplacement des données manquantes est à même d'apporter des preuves à l'abri des biais.
- L'analyse critique d'une information thérapeutique porte tout autant sur la fiabilité de cette information (validité méthodologique et statistique) que sur sa pertinence clinique.

Un biais survient lorsqu'il existe une différence entre les groupes dans un facteur influençant la valeur du critère de jugement.

Un biais de confusion survient quand il existe une différence, entre les deux groupes, au niveau de facteurs pouvant être confondue avec l'effet du traitement étudié au niveau du critère de jugement : différence de pronostic initial des patients, différence dans la prise en charge des patients...

Un biais de mesure survient quand il existe une différence de « méthode » de mesure du critère de jugement entre les groupes.

En pratique, il est très difficile d'objectiver la présence d'un biais, car rien ne garantit que les données montrant le biais aient été recueillies ou rapportées. De ce fait, des principes méthodologiques qui empêchent la survenue des biais ont été développés. Par exemple, le groupe contrôle évite la survenue d'un biais de confusion. Lorsqu'un essai applique correctement ces principes, la survenue d'un biais devient impossible. Seule persiste la possibilité d'une erreur aléatoire, qui sera gérée par la recherche d'une différence statistiquement significative, le calcul a priori du nombre de sujets nécessaires pour garantir une puissance

Tableau Échelle de niveau de preuve pour l'évaluation thérapeutique de la Haute Autorité de santé française*

Niveau 1	- Essais comparatifs randomisés de forte puissance - Méta-analyse d'essais comparatifs randomisés - Analyse de décision basée sur les études bien menées	Preuve scientifique établie
Niveau 2	- Essais comparatifs randomisés de faible puissance - Études comparatives non randomisées bien menées - Études de cohorte	Présomption scientifique
Niveau 3	- Études cas-témoins - Études comparatives comportant des biais importants - Études rétrospectives	Faible niveau de preuve scientifique
Niveau 4	- Série de cas - Études épidémiologiques descriptives	

* Le niveau 1 représente le plus haut niveau de preuve.

statistique élevée et par la répétition des études. Les erreurs aléatoires s'annulant en moyenne, le résultat d'une méta-analyse est soumis à une erreur aléatoire réduite par rapport au résultat d'une étude prise isolément.

Randomisation

La randomisation évite le biais de sélection qui survient lorsqu'il y a sélection des patients recevant le traitement évalué (sélection des patients au meilleur pronostic par exemple). L'allocation aléatoire des traitements fait que la nature du traitement reçu dans l'essai par un patient est forcément indépendante de son état et de son devenir, évitant ainsi tout processus de sélection pouvant introduire ce biais. C'est plus par son imprévisibilité que par son caractère aléatoire que la randomisation protège contre le biais de sélection. Ainsi, une randomisation par enveloppe dans un essai en ouvert ne met pas à l'abri d'un biais de sélection, car la nature du traitement que recevra un patient est prévisible avant son inclusion.

Double aveugle

Le double aveugle évite les biais de suivi et de mesure. Le biais de suivi est produit par une prise en charge différente, en fonction de la nature du traitement reçu, des patients durant l'étude (au niveau par exemple des traitements concomitants ou de la prise en charge médicale en général). Le biais de mesure est produit par une mesure différente du critère de jugement en fonction du traitement reçu (sous-estimation de la fréquence d'un événement influencée par la connaissance de la nature du traitement reçu par le patient par exemple). Le double insu, en rendant impossible la connaissance du traitement reçu, évite ces différences d'attitude entre les deux groupes dans la prise en charge et la mesure du critère de jugement. Lorsque le double aveugle est impossible, comme souvent dans un essai comparant traitement médical et traitement chirurgical, l'évaluation en aveugle du critère de jugement permet d'éviter le biais de mesure mais pas le biais de suivi.

Analyse en intention de traiter

L'analyse en intention de traiter et le remplacement des données manquantes par une méthode conservatrice évite le biais d'attrition. L'attrition correspond à une diminution du nombre de patients analysés par rapport au nombre de patients randomisés en raison de patients sortis de l'étude ou perdus de vue. Cette attrition est susceptible d'entraîner un biais si elle n'est pas indépendante du devenir des patients ou du traitement reçu. Pour éviter ce biais, tous les patients inclus doivent être analysés dans le groupe où ils ont été randomisés (analyse en intention de traiter). Le maintien dans l'analyse des perdus de vue, chez qui par définition la valeur du critère de jugement est inconnue, nécessite de remplacer cette valeur manquante par une valeur arbitraire. Le choix de la valeur de remplacement ne doit pas favoriser le traitement étudié et peut être la dernière valeur connue pour un critère continu ou la survenue d'un échec du traitement pour une variable binaire (analyse suivant le scénario du pire pour le traitement étudié).

MÉTA-ANALYSE

La méta-analyse effectue la synthèse de tous les essais thérapeutiques répondant à la même question thérapeutique. À partir des données de chaque étude, la méta-analyse donne la meilleure estimation possible de l'efficacité et de la sécurité. La méthodologie des méta-analyses garantit que la synthèse sera rigoureuse et non arbitraire.

Les principes de la méta-analyse sont :

- de rechercher tous les essais publiés et non publiés apportant une réponse à la question de la méta-analyse (par exemple, l'efficacité sur les pathologies cardiovasculaires des diurétiques dans l'hypertension artérielle). La recherche des essais non publiés permet d'éviter le biais de publication, lié au fait que les essais négatifs sont moins facilement publiés que les essais positifs ;
- de ne retenir que les essais de meilleure qualité afin d'éviter de biaiser la méta-analyse par l'introduction d'études elles-mêmes biaisées ;
- d'estimer l'efficacité à partir des résultats de chaque essai en calculant par exemple un risque relatif afin d'éviter de raisonner sur la signification statistique de chaque étude qui pose des problèmes de multiplicité des comparaisons.

Ces principes permettent d'éviter les travers des synthèses non systématiques comme la revue générale de la littérature où, assez fréquemment, ne sont rapportés que les résultats en accord avec l'opinion exprimée, les autres essais étant passés sous silence.

Pour être exhaustive, une méta-analyse doit avoir recherché les études dans au moins 2 bases bibliographiques comme PubMed. La recherche des essais non publiés fait appel maintenant aux registres d'essais cliniques où les essais sont déclarés au moment de leur mise en place.

CRITIQUER ET INTERPRÉTER UNE INFORMATION THÉRAPEUTIQUE

La critique d'une information thérapeutique (revendication d'efficacité) a pour objectif de déterminer si celle-ci repose sur un résultat fiable et cliniquement pertinent. Il s'agit de vérifier que des preuves solides existent en faveur d'une efficacité cliniquement pertinente. Si de telles preuves existent, le traitement évalué pourra être rajouté à l'arsenal thérapeutique de la pathologie considérée. Si les résultats disponibles présentent des limites méthodologiques et/ou manquent de pertinence clinique, le traitement ne peut pas être utilisé en pratique courante et restera dans le domaine de l'évaluation afin de parfaire la recherche de preuves.

Pour cette lecture critique, il convient de s'assurer de :

- la fiabilité méthodologique du résultat, en vérifiant que l'essai est à l'abri des biais : randomisation imprévisible, double aveugle, absence d'attrition assurée par une analyse en intention de traiter et l'absence de perdus de vue ou le remplacement des données manquantes ;

– la réalité statistique du résultat, en vérifiant que la différence entre les 2 groupes est statistiquement significative au seuil habituel de 5 %, et qu'il n'y a pas de problème de multiplicité des tests statistiques comme une absence de critère de jugement principal, l'obtention du résultat dans une analyse en sous-groupe, ou la recherche répétée au cours du temps de l'effet du traitement. En d'autres termes, la réalité statistique de la différence observée est assurée seulement si une différence statistiquement significative existe sur le critère de jugement principal avec l'ensemble des patients inclus et à un temps de mesure unique défini a priori ;

– la pertinence clinique du résultat, en vérifiant que le critère de jugement est un critère clinique correspondant directement aux problèmes médicaux des patients, que les patients sont représentatifs des patients pris en charge en pratique courante, que la taille de l'effet est connue avec une bonne précision (intervalle de confiance à 95 % étroit) et qu'elle est suffisamment importante pour être intéressante en pratique, surtout compte tenu des effets indésirables (c'est-à-dire que la balance bénéfice-risque est favorable). En cas d'essais versus traitement actif, le traitement contrôle doit aussi être pertinent (meilleur traitement disponible). ■

L'auteur déclare n'avoir aucun conflit d'intérêts concernant les données publiées dans cet article.

Pour en savoir plus

► Lecture critique et interprétation des résultats des essais cliniques pour la pratique médicale

Cucherat M, Lièvre M, Leizorovicz A, Boissel JP
Paris : Flammarion Médecine-Sciences, 2004

MINI TEST DE LECTURE

VRAI ou FAUX ?

- 1 La pertinence clinique d'un résultat dépend de son niveau de signification statistique.
- 2 Pour éviter les biais, il convient d'exclure de l'analyse les patients mauvais observants.
- 3 La randomisation évite le biais de sélection.
- 4 Dans une situation où le double aveugle est impossible, l'évaluation du critère de jugement en aveugle du traitement reçu évite tous les biais.

1) Pertinence clinique et signification statistique sont deux concepts différents. La signification statistique s'intéresse à la réalité de l'effet compte tenu du risque d'erreur statistique alpha. La pertinence clinique concerne l'intérêt médical et dépend de la pertinence du critère de jugement utilisé, de la taille de l'effet et de la balance bénéfice-risque.

2) L'exclusion de patients randomisés entraîne une attrition et empêche de faire l'analyse en intention de traiter. L'essai devient exposé au biais d'attrition.

3) La randomisation évite qu'il soit possible de sélectionner les patients inclus dans le groupe traité, en fonction de leur pronostic par exemple. C'est un élément clé de la protection contre les biais.

4) Seul le biais de mesure est évité, mais pas le biais de suivi. En raison de l'absence d'aveugle, rien n'exclut que les patients aient été pris en charge de manière différente suivant le groupe dans lequel ils étaient.

Réponses : F, F, V, F

MINI TEST DE LECTURE de la question 314, page 783

A / VRAI ou FAUX ?

- 1 Les exanthèmes sont toujours d'origine infectieuse.
- 2 L'exanthème lié à la primo-infection VIH est une urgence diagnostique et thérapeutique.
- 3 Un exanthème survenant 12 h après le début d'un traitement médicamenteux évoque une toxidermie.
- 4 Le diagnostic du syndrome de Kawasaki est biologique.

B / VRAI ou FAUX ?

- 1 Une érythrodermie est une éruption transitoire et fugace.
- 2 L'érythrodermie est une urgence dermatologique et justifie une hospitalisation.
- 3 L'eczéma et le psoriasis sont des causes possibles d'érythrodermie.

- 4 L'apparition d'une éruption cutanée après la prise d'amoxicilline dans un contexte de mononucléose infectieuse contre-indique la reprise de la pénicilline.

C / QCM

Parmi ces éléments, quels sont les arguments pour un syndrome de Kawasaki ?

- 1 Survenue chez un adulte.
- 2 Fièvre élevée > 38,5 °C depuis plus de 5 jours.
- 3 Antécédent d'eczéma.
- 4 Anévrismes coronariens à l'échographie cardiaque.
- 5 Syndrome inflammatoire biologique avec thrombocytose.

Réponses : A : F, V, F, F, F / B : F, V, V, F, F / C : 2, 4, 5.